❏   1027

A new approach Based on Quantum Clustering and Wavelet Transform for breast cancer Classification: Comparative study

Nezha Hamdi[1], Khalid Auhmani[2], Moha. M'Rabet Hassani[1]
1 Department of Physics, Faculty of sciences Semlalia, Cadi Ayyad University, Marrakech, Morocco
Nezha_hamdi@yahoo.com
2 Department of Industrial Engineering, National school of applied sciences, Cadi Ayyad University, Safi, Morocco
k.auhmani@uca.ma

Article Info

ABSTRACT

Feature selection involves identifying a subset of the most useful features that produce the same results as the original set of features. In this paper, we present a new approach for improving classification accuracy. This approach is based on quantum clustering for feature subset selection and wavelet transform for features extraction. The feature selection is performed in three steps. First the mammographic image undergoes a wavelet transform then some features are extracted. In the second step the original feature space is partitioned in clusters in order to group similar features. This operation is performed using the Quantum Clustering algorithm. The third step deals with the selection of a representative feature for each cluster. This selection is based on similarity measures such as the correlation coefficient (CC) and the mutual information (MI). The feature which maximizes this information (CC or MI) is chosen by the algorithm. This approach is applied for breast cancer classification. The K-nearest neighbors (KNN) classifier is used to achieve the classification. We have presented classification accuracy versus feature type, wavelet transform and K neighbors in the KNN classifier. An accuracy of 100% was reached in some cases.

Corresponding Author:

Khalid Auhmani
Departement of Industrial Engineering,
National School of Applied Sciences,
Cadi Ayyad University,
BP 63 Safi, Morocco
Email: k.auhmani@uca.ma

1.   INTRODUCTION

The high dimensional nature of many data in bioinformatics, has given rise to a wealth of feature subset selection techniques. Feature selection aims identifying a subset of the most useful features that allows the same results as the original set of features. Feature subset selection is an effective method for removing irrelevant features, improving learning accuracy, and improving classification accuracy. Many methods have been studied for different applications. They are generally classified into three categories: the Wrapper, Filter, and Embedded [1].

The Wrapper approaches use the classification error rate as a evaluation criteria [2]. They then incorporate the classification algorithm in the search and selection of attributes. These methods allow the obtaining of high performance. However, the use of such methods requires for each subspace of attributes to perform classification, which can become costly in calculation time especially when the dimension d of the input space is large. These methods are very dependent of the used classification algorithm.

Filter approaches use an evaluation function based on the characteristics of all data, independently of any classification algorithm [3-7]. These methods are fast, general and less expensive in computation time,

which allows them to operate more easily with databases of very large dimensions. However, as they are independent of the classification stage, they do not guarantee to reach the best classification accuracy.

In order to combine the advantages of both methods, hybrid algorithms "embedded" have been proposed. The feature selection process is performed in conjunction with the classification process. A filter-type evaluation function is first used to screen the most discriminating feature subspace. Then the error rates of misclassification, by considering each discriminant subspace previously selected, are compared in order to determine the final subspace [8, 9].

Due to their computational efficiency and independence of any classification algorithm, the "filter" approaches are more popular and commonly used.

With respect to the filter feature selection methods, the application of cluster analysis has been demonstrated to be more effective than traditional feature selection algorithms [1].

In contrary with feature selection for supervised learning systems, relatively few approaches have been proposed for unsupervised learning (automatic classification or clustering). Indeed, the feature selection problem for automatic classification is a much more difficult problem in supervised cases (discrimination) where the data are labeled [10]. Another important problem associated with classification concerns the automatic determination of the number of clusters that is clearly influenced by the outcome of the feature selection.

In cluster analysis data are divided into groups (clusters). The goal is that the objects within a group be similar (or related) to one another and different from (or unrelated to) the objects in other groups. This step is useful for feature selection purposes.

In our study we propose a new method for feature selection based on clustering. This method works in two steps. In the first features are grouped in clusters. We apply the quantum clustering algorithm (QC) presented in [11] for this purpose. In the second step, the most representative feature that is strongly related to target classes is selected from each cluster to form the final subset of features. This selection is performed by using the similarity measures such as correlation coefficient (CC) and the mutual information (MI). The feature which maximizes this information is chosen by the algorithm.

2.   MATERIALS AND METHODS

2.1 Framework

Mammography is currently the most effective imaging modality for breast cancer screening. However, the sensitivity of mammography is highly challenged by the presence of dense breast parenchyma, which deteriorates both detection and characterization tasks [12-13]. Computer Aided diagnosis (CAD) systems have been developed to aid radiologists in detecting mammographic lesions, characterized by promising performance [14-15]. CAD systems for aiding the decision concerning biopsy and follow-up are still under development.

Various CAD diagnosis algorithms have been proposed. These algorithms are based on extracting image features from regions of interest (ROIs) and estimating the probability of malignancy. A variety of computer-extracted features and classification schemes have been used to automatically discriminate between benign and malignant image. The majority of these studies have followed three steps.

The first step deals with two problems: suppress the noise and enhance the contrast between the region of interest (ROI) and background. It is the task of preprocessing. The second step deals with the extraction of features. This task can be done on the image without any transformation or on the image after an adequate transformation. The third step deals with the selection of features and classification.

Figure 1 presents the framework we realized to improve the classification accuracy of mammoghraphic images through the proposed method.

The proposed algorithm was tested on the original mammographic images of MIAS database [16].

Preprocessing is an important issue in low-level image processing. The underlying principle of preprocessing is to enlarge the intensity difference between objects and background and to produce reliable representations of breast tissue structures. More details of the proposed preprocessing method can be found in [17, 18].

Before extracting some features we decided to transform the input image by a wavelet transform. This choice is imposed by the fact that microcalcifications are characterized by high frequency behavior. Microcalcifications are a sign of possible presence of cancer in breast.

The basic idea behind wavelet transform is to analyze different frequencies of an image using different scales. High frequencies are analyzed using low scales whilst low frequencies are analyzed in high scales. This is a far more flexible approach than the Fourier transform, enabling analysis of both local and global features [19].

In this work we use the discrete wavelet transform (DWT) based on the concept of multiresolution analysis (MRA) introduced by Mallat [20] and the Double Density Wavelet transform (DDWT) introduced by Ivan Selesnick [21, 22]. Images are decomposed in three levels then features are extracted.
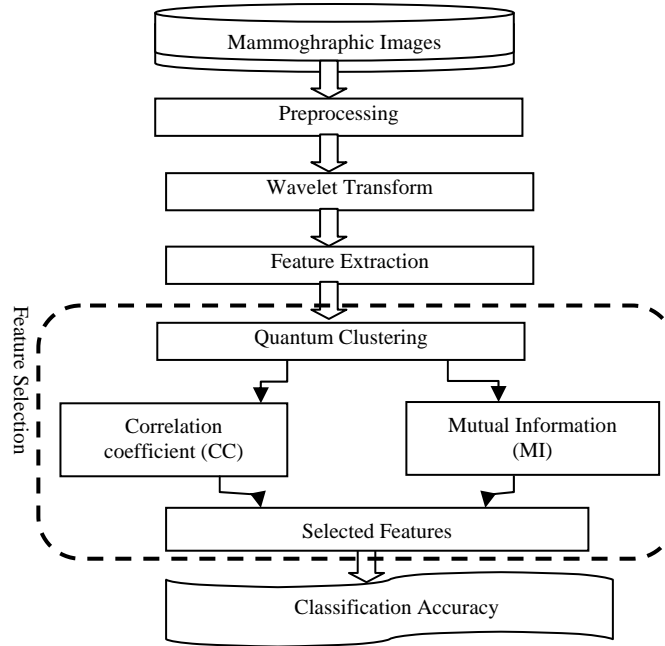

Figure 1. Framework of the proposed feature subset selection algorithm

2.2 Discrete Wavelet Transform

For the DWT, an image is first transformed by a filter in the horizontal direction. The high-pass filter and the low-pass filter are finite impulse response filters. The filtered outputs are down sampled by a factor of 2 in horizontal direction. The signals are then filtered by an identical filter in the vertical direction and down sampled by a factor of 2 again. The result is a decomposition of the image into four sub-bands. An approximation sub-band and three details sub-bands.

The three details sub-bands represent the finest scale wavelet coefficients while the approximation sub-band corresponds to coarse level coefficients. To obtain the next coarse level of wavelet coefficients, the approximation sub-band is further decomposed and critically sampled. This results in two-level wavelet decomposition as shown in Figure 2.
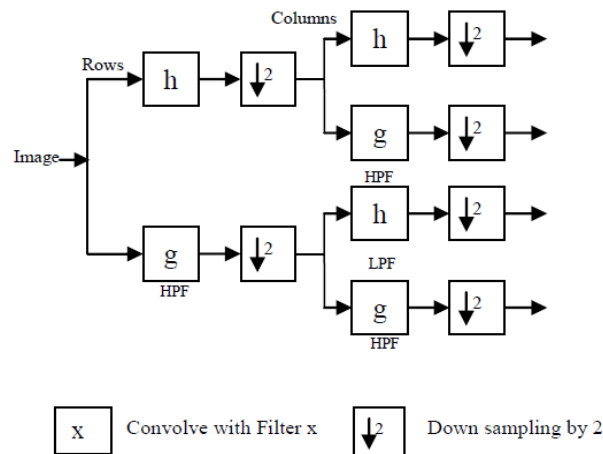

Figure 2. Discrete wavelet transform Filter Bank for One-Level Image Decomposition [23]

2.3 Double Density Wavelet Transform

Although the DWT is a powerful signal processing tool, it has two severe disadvantages [24]:

a. Lack of shift-invariance, which means that minor shifts in the input signal, can cause major variations in the distribution of energy between wavelet coefficients at different scales.

b. Since the wavelet filters are separable and real, it causes poor directional selectivity for diagonal features

The double density wavelet transform provides higher directional selectivity, better peak signal to noise ratio and visual perception than the spatial domain methods and other frequency domain methods [23]. DDWT is a shift insensitive, directional, complex wavelet transform with a very low redundancy factor of two regardless of the number of scales. Double density wavelets have a single scaling function and two wavelet functions. Figure 3 shows the DDWT scheme [23].
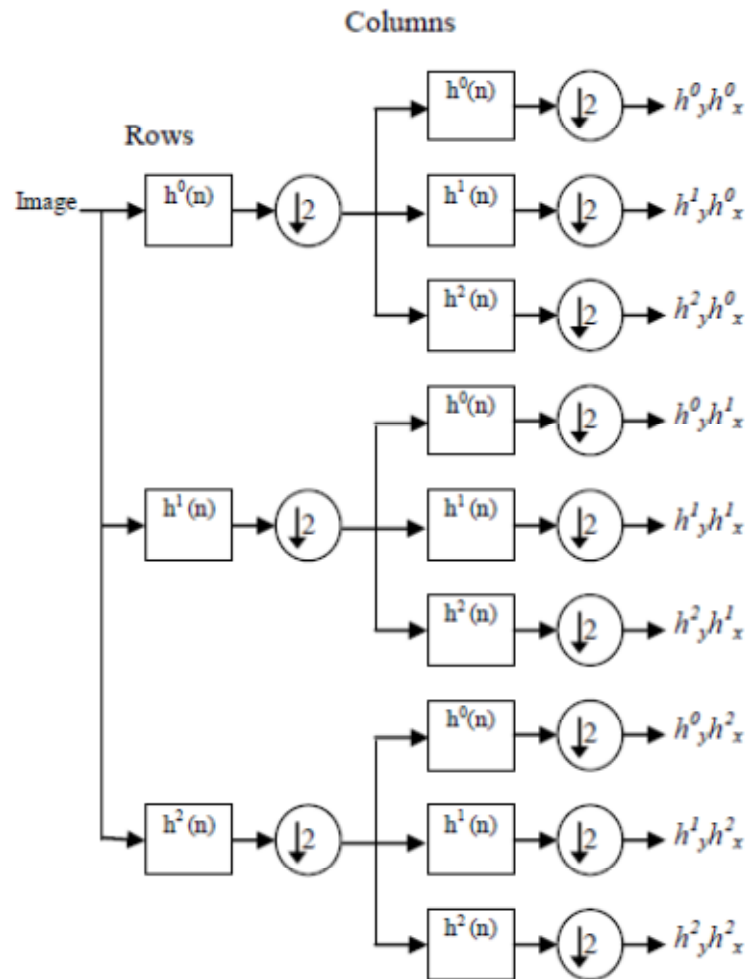

Figure 3. Double density wavelet transform scheme [23]

2.4 Features Extraction

Features are extracted from a set of two classes labeled images (normal and abnormal). The following features are extracted:

- A vector of 24 texture descriptors is formed from a multi-level histogram of 3, 5, 7, and 9 bins [25].

- A vector of descriptors is calculated from the first order of statistical moments [26].

- Three of the six parameters introduced by Tamura are used [27, 28].

- Radon's characteristics are calculated [29, 30].

- Zernike's moments of order $n = 12$ are calculated corresponding to 49 features [31].

If the used wavelet transform is the double density discrete wavelet transform (DDWT), the image is divided into nine sub-images, so the number of features is $9 * 139 = 1251$. If the transformation is based on the

discrete wavelet transform (DWT), the original image is decomposed into four sub-images, and the number of calculated features is 4*139= 556. Each original image will be represented by a group of 556 or 1251 features.

2.5 Feature Selection

Extra features can increase computation time, and can impact the accuracy of the Detection System. Feature selection improves classification by searching for the subset of features, which best classify the training data [32]. Feature selection reduces the dimensionality of data by selecting only a subset of measured features (predictor variables) to create a model [33].

Feature subset selection is a process of identifying and removing irrelevant and redundant features. Irrelevant features, along with redundant features, severely affect the accuracy of the learning machines [34, 35]. Thus, feature subset selection should be able to identify and remove as much of the irrelevant and redundant information as possible

Many feature subset selection algorithms have been developed. Some of theme can effectively eliminate irrelevant features but fail to handle redundant features, yet some of others can eliminate the irrelevant while taking care of the redundant features [1,35]. We propose a method that falls into the second group.

Clustering occurs in unsupervised learning tasks where the labels of training examples are not known a priori. The goal is to find an organization point cloud corresponding to the training examples, M areas, called clusters. We will use clustering in a particular context in which it will be applied directly to vectors of attributes: we will see that we need to reorganize these attributed by grouping those that are "closer" to each other in clusters (classes), forming superclasses. We will use a particular clustering approach: quantum clustering

Quantum Clustering

Clustering occurs in unsupervised learning tasks where the labels of training examples are not known. The goal is to find an organization point cloud corresponding to the training examples, M areas, called clusters. The result is a forest and each tree in the forest represents a cluster [1]. We will use clustering in a particular context in which it will be applied directly to feature vector. We will see that we need to reorganize these features by grouping those that are "closer" to each other in clusters. We will use a particular clustering approach: quantum clustering.

Horn [11] estimated locally probability densities at the point x by observing the training set around this point. They are implemented by the windows Parzen technique [36]

Bravais-Pearson Correlation Coefficient (CC)

The correlation coefficient of Bravais-Pearson is a statistical index that expresses the intensity and direction (positive or negative) of the linear relationship between two quantitative variables. It is a measure of the linear link, i.e. the ability to predict a variable x by another variable y using a linear model.

It allows measuring the relationship intensity between two quantitative variables. It is therefore an important parameter in the linear regressions (single or multiple) analysis. However, this coefficient is zero (r = 0) when there is no linear relationship between the variables (which does not exclude the existence of a non linear relationship). Moreover, the coefficient has a positive sign if the relationship is positive (direct, increasing) and negative sign if the relationship is negative (inverse, decreasing).

Mutual Information (MI)

The mutual information of a pair of variables (X, Y) represents the dependence degree in the probabilistic sense. It measures the amount of information brought by a random variable on another. It is a reduction of uncertainty about a random variable due to the knowledge of another.

The mutual information and the correlation coefficient procedures are not based on a particular model. The construction of a K-nearest neighbors (KNN) model is feasible on the d selected variables. The obtained results for the QC-CC and QC-IM model will be presented.

3. RESULTS AND DISCUSSIONS

In order to measure the performance of the QC-FS algorithm we have used the KNN classifier. Experiments are performed under Matlab software. The experiment carried out deals with the effect of wavelet transform in association with similarity measure on the classification accuracy. Figures 4 and 5 show results corresponding to double density and discreet wavelet transforms associated with the correlation coefficient and mutual information respectively.

First, we note the superiority of DDWT independently of used features and employed similarity measure for feature selection. Indeed, the best classification accuracies are obtained with the Zernike moments (100%) followed by the statistical moments (91.5 %). However, associated with the discrete wavelet transform, Zernike moments give the least satisfactory results, although they are the most successful with DDWT. Here, we show a relation between the feature type and the used Transformation. For discrete wavelet (DWT) we reach low performance (61.6%).

Concerning the contribution of CC and MI similarity measures in classification accuracy we can deduce that MI provides results better than CC. Indeed we find that the average of classification accuracy for MI is about 99%, while this average is around 85% for CC.
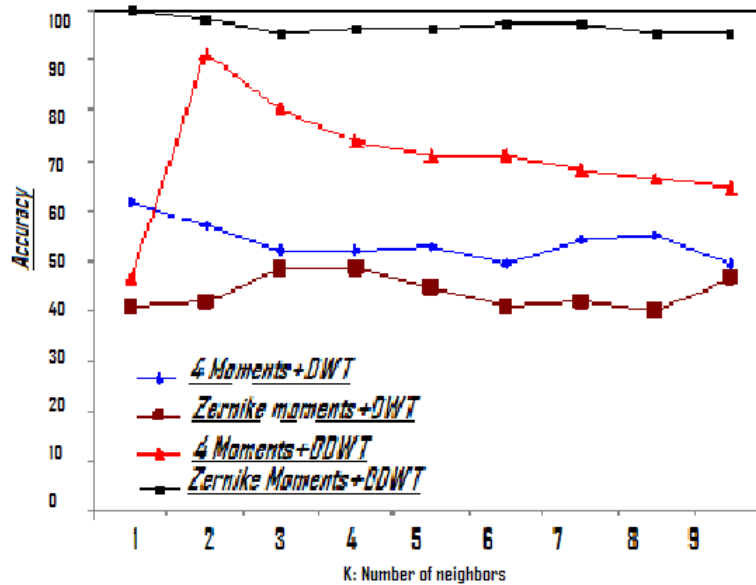

Figure 4. Classification accuracy versus wavelet transform and features types associated with similarity measure-based Corelation coefficient (CC)
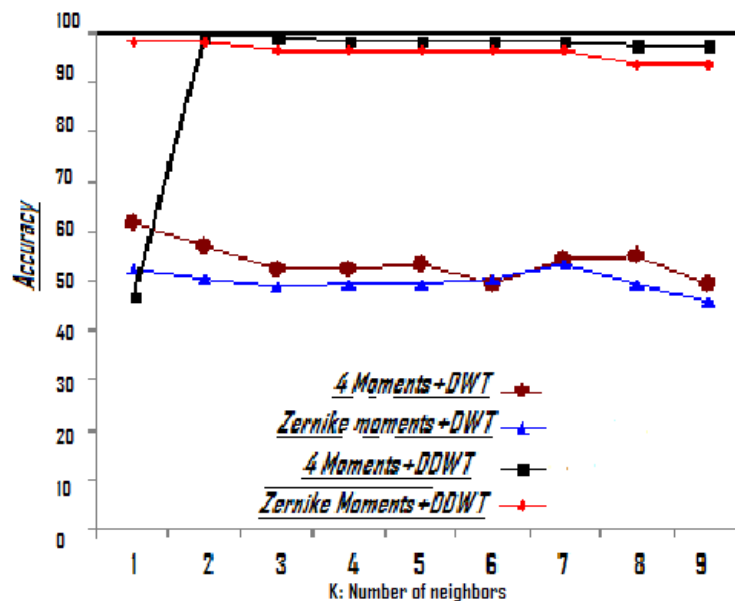

Figure 5. Classification accuracy versus wavelet transform and features types associeted with similarity measure-based Mutual Information (MI)

4. CONCLUSION

In this paper we have presented a new approach to improve classification accuracy. This approach is based first on wavelet transform for feature extraction and second on quantum clustering to select relevant features. We have applied a double density wavelet transform on the mammographic image and then we extracted some features to be used in classification. The subset feature selection algorithm achieves the selection task in two steps. The first step deals with the clustering of data. This clustering is carried out by the quantum clustering algorithm. The second step deals with the selection of a representative feature of each cluster. This selection is based on mutual information or on correlation coefficient.

Classification accuracy of mammographic image was computed versus the feature type and the used wavelet transform.

We noticed that the DDWT provides good accuracies independently of the used features and the employed similarity measure. Indeed, the best classification accuracies were reached with the Zernike moments (100%) followed by the statistical moments (91.5 %). However, associated with the discrete wavelet transform, Zernike moments give the least satisfactory results, although they are the most successful with DDWT. For discrete wavelet (DWT) we reach low performance (61.6%).

Concerning the contribution of CC and MI similarity measures in classification accuracy we can deduce that MI provides results better than CC. Indeed we find that the average of classification accuracy for MI is about 99%, while this average is around 85% for CC.

REFERENCES

[1] Qinbao Song, Jingjie Ni and Guangtao Wang; A Fast Clustering-Based Feature Subset Selection Algorithm for High-Dimensional Data; IEEE Transactions on Knowledge and Data Engineering, (Volume: 25, Issue: 1), 2011.
[2] R. Kohavi and G. John. Wrappers for feature subset selection. Artificial Intelligence, pages 273–324, 1997.
[3] Dash M. and Liu H., Feature Selection for Classification, Intelligent Data Analysis, 1(3), pp 131-156, 1997.
[4] Langley P., Selection of relevant features in machine learning, In Proceedings of the AAAI Fall Symposium on Relevance, pp 1-5, 1994.
[5] Souza J., Feature selection with a general hybrid algorithm, Ph.D, University of Ottawa, Ottawa, Ontario, Canada, 2004.
[6] X. He, D. Cai, and P. Niyogi. Laplacian score for feature selection. In Proceedings of the Advances in Neural Information Processing Systems 'NIPS 05', pages 507–514, Vancouver, Canada, December 2005.
[7] L. Talavera. Feature selection as a preprocessing step for hierarchical clustering. In Proceedings of the 16th International Conference on Machine Learning 'ICML 99', pages 433–443, Bled, Slovenia, 1999.
[8] S. Das. Filters, wrappers and a boosting-based hybrid for feature selection. In Proceedings of the 18th International Conference on Machine Learning 'ICML 01', pages 74–81, Williamstown, MA, USA, June 2001.
[9] Guyon I. and Elisseeff A., An introduction to variable and feature selection, Journal of Machine Learning Research, 3, pp 1157-1182, 2003.
[10] G Guyon, I., S. Gunn, M. Nikravesh, et L. Zadeh (2006). Feature Extraction, Foundations and Applications, Editors. Series Studies in Fuzziness and Soft Computing, Physica-Verlag.Springer.
[11] H. David ; GOTTLIEB Assaf, Algorithm for data clustering in pattern recognition problems based on quantum mechanics, Physical review letters , 2002, vol. 88, no1, pp. 018702.1-018702.4
[12] P.M. Sampat, M.K. Markey, and A. C. Bovik, "Computer-aided detection and diagnosis in mammography" in Handbook of Image and Video Processing, 2nd ed., A. C. Bovik Ed. Academic Press, 2005, pp. 1195-1217.
[13] D.D. Adler and M.A. Helvie, "Mammographic biopsy recommendations", Curr. Opin. Radiol., vol 4., pp. 123-129, 1992
[14] M.L. Giger, N. Karssemeijer and S.G. Armato, "Computer-aided diagnosis in medical imaging", IEEE Trans. on Med. Imaging, vol. 20, pp. 1205-1208, 2001.
[15] C.J. Vyborny, M.L. Giger and R.M. Nishikawa, "Computer-aided detection and diagnosis of breast cancer", Radiologic Clinics of North America, vol. 38, pp. 725-740, 2000.
[16] http://peipa.essex.ac.uk/info/mias.html
[17] N. Hamdi, K. Auhmani and M.M. Hassani; Design of a High-Accuracy Classifier Based on Fisher Discriminant Analysis: Application to Computer-Aided Diagnosis of Microcalcifications; Proceeding of The 2008 International Conference on Computational Science and Its Applications, Perugia Italy; IEEE - Computer Society, pp. 267-273, 2008.
[18] N. Hamdi, K. Auhmani, M.M. Hassani, Computer Aided Diagnosis: Automated detection and enhancement of microcalcifications in digitized mammograms using wavelet decomposition and local gray thresholding, PCN Journal, 47, 75-78, May 2009.
[19] Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J.M.: Wavelet Toolbox User's Guide. MathWorks Inc., Massachusetts, (1996).
[20] S.G. Mallat, A Theory for Multiresolution Signal Decomposition: The Wavelet Representation, IEEE Transactions on Pattern Analysis and Machine Intelligenc, Volume 11 Issue 7, July 1989, Page 674-693.
[21] W. Selesnick. The double-density dual-tree DWT. IEEE Trans. on Signal Processing, 52(5):1304-1314, May 2004.
[22] W. Selesnick. The double density dwt. In A. Petrosian and F. G. Meyer,editors, Wavelets in Signal and Image Analysis: From Theory to Practice. Kluwer.

[23] S. Arivazhagan, L. Ganesan and C.N. Savithri, "Effective multi-resolution transform identification for characterization and classification of texture groups", ictact journal on image and video processing, november 2011, volume: 02, issue: 02.

[24] Varun P. Gopi, V. Suresh Babu, Dilna C, "Image Resolution Enhancement Using Undecimated Double Density Wavelet Transform", Signal Processing: An International Journal (SPIJ), Volume (8) : Issue (5) : 2014

[25] Hadjidemitriou, E., Grossberg, M.D., & Nayar, S.K. Multiresolution histograms and their use for recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26(7), 831-847. (2004)

[26] S. HERLIDOU, Caractérisation tissulaire en IRM par l'analyse de texture. Étude du tissu musculaire et de tumeurs intracrâniennes, université de Renne1, 1999.

[27] H. TAMURA, S. MORI et T. YAMAWAKI. Texture features corresponding to visual perception, IEEE Transactions on Systems, Man and Cybernetics, SMC-8(6): 460–473, 1978.

[28] P. HOWARTH et S. RÜGER. Evaluation of texture features for content-based image retrieval. In Proceedings of the International Conference on Image and Video Retrieval (CIVR'04), volume LNCS 3115, pages 326–334, Dublin, Ireland, jul 2004.

[29] Deans, S.R. Hough Transform from the Radon Transform, IEEE Trans. On Patt. Anal. and Mach. Intell., Vol. PAMI-3, No. 2, pp. 185_188, 1981.

[30] Murphy, L.M. Linear feature detection and enhancement in noisy images via the Radon transform, Pattern Recognition Letters, No. 4, pp. 279-284, 1986.

[31] C.W. Chong, P. Raveendran, and R. Mukun-dan. A comparative analysis of algorithms for fast computation of zernike moment. Pattern Recognition, 36:731-742, 2003.

[32] Thuzar Hlaing, Feature Selection and Fuzzy Decision Tree for Network Intrusion Detection, International Journal of Informatics and Communication Technology (IJ-ICT) Vol.1, No.2, pp. 109~118, December 2012

[33] Lichen Xun, Gang Zheng, ECG Signal Feature Selection for Emotion Recognition, TELKOMNIKA, Vol. 11, No. 3, pp. 1363 ~ 1370, March 2013

[34] Hall M.A., Correlation-Based Feature Selection for Discrete and Numeric Class Machine Learning, In Proceedings of 17th International Conference on Machine Learning, pp 359-366, 2000.

[35] Kohavi R. and John G.H., Wrappers for feature subset selection, Artif. Intell., 97(1-2), pp 273-324, 1997.

[36] R.O. Duda, P.E. Hart and D.G. Stork. Pattern Classification. Wiley-Interscience, 2nd ed., 2001.